

Deprem Öncülü İyonküresel Sinyallerin K-En Yakın Komşular Algoritması ile Sınıflandırılması

Ali Çınar, Seçil Karatay, Cafer Budak*, Faruk Erken
Kastamonu Üniversitesi
Elektrik Elektronik Mühendisliği Bölümü
Kastamonu

acinar@kastamonu.edu.tr, skaratay@kastamonu.edu.tr, ferken@kastamonu.edu.tr

*Dicle Üniversitesi
Biyomedikal Mühendisliği Bölümü
Diyarbakır
cafer.budak@dicle.edu.tr

Özet: Bu çalışmada, bir depremin iyonkürede meydana getirdiği öncül sinyallerin sınıflandırılması için, K-En Yakın Komşular (KNN) algoritması, iyonküresel Toplam Elektron İçeriği (TEİ) verilerine uygulanmıştır. KNN algoritmasının, 24 deprem verisinin 22 tanesini doğru, 2 tanesi ise hatalı olarak sınıflandırdığı gözlenmiştir.

Abstract: In this study, the K-Nearest Neighbors (KNN) algorithm is applied to ionospheric Total Electron Content (TEC) to classify the precursor signals due to an earthquake in the ionosphere. It is observed that the KNN algorithm classified 22 of 24 earthquake data as correct and 2 as incorrect.

1. Giriş

İyonküre, güneş ve kozmik ışınlarla iyonlaşmış Yer atmosferinin bir katmanıdır ve Yer yüzeyinden yaklaşık 50 ila 1000 km yükseklikler arasında yer alır. İyonküreyi en iyi tanımlayan parametre, konum, zaman, mevsimler, güneş, jeomanyetik ve sismik hareketlilikle önemli değişimler sergileyen elektron yoğunluğudur. Atmosferik dalgalar iyonküreye ulaştığında, iyonküredeki elektronların yoğunluğunda saptanabilir değişikliklere neden olurlar. Yerküresel Konumlama Sistemi (YKS) gibi Yerküresel Navigasyon Uydu Sistemleri'nden (GNSS-Global Navigation Satellite System) gelen sinyaller iyonküreden geçerken bu değişiklikler kaydedilebilmekte ve ölçülebilmektedir. İyonkürede meydana gelen bu değişimlerin ölçüsü, iyonkürenin Toplam Elektron İçeriği (TEİ) ile ifade edilebilmektedir. 1 metrekare kesitli bir silindir boyunca toplam serbest elektron miktarı olarak da yorumlanan TEİ'nin birimi TECU'dur. 1 TECU, metrekarede 10^{16} elektrona eşittir [1, 2].

Son yıllarda, iyonküre ve deprem ilişkisi üzerine yoğun çalışmalar yapılmıştır [3, 4]. Bu çalışmalarda, büyük depremlerden önce ve depremler sırasında iyonkürenin, sismik hareketliliğin etkisiyle dalga benzeri salınımlar olarak ortaya çıkan bazı bozulmalar ve aykırılıklar (anomali) sergilediği tespit edilmiştir. Özellikle büyük depremlerden günler önce uydulardan, TEİ seviyelerinde bir zirve de dâhil olmak üzere istatistiksel olarak önemli değişiklikler kaydedilmiştir. Deprem kaynaklı iyonküresel bozulmaların ve depremlerden önce meydana gelen aykırı sinyallerin tespit ve tahmini için birçok model geliştirilmiş ve uygulanmıştır [5, 6]. İyonkürenin TEİ-deprem izleme ile gerçekleştirilen bu modellerde, TEİ değişimlerinin sürekli izlenip analiz edilerek, deprem tahmininde bulunulmasını sağlamaktadır.

Bu çalışmada, YKS-TEİ verilerine ve günlük jeomanyetik indislere dayalı olarak deprem öncüllerinin sınıflandırılması için makine öğrenmesi tabanlı olan K-En Yakın Komşular (KNN-K Nearest Neighbors) algoritması uygulanmıştır. KNN algoritmasının performansı, 11 Ağustos 1999 tam güneş tutulması ve 17 Ağustos 1999 Marmara depremini kapsayan iki vaka için yirmi günlük bir zaman aralığında test edilmiştir. Algoritma ile ilgili bilgilere Bölüm 2'de, elde edilen bulgulara da Bölüm 3'te yer verilmiştir.

2. K-En Yakın Komşular Algoritması

Herhangi bir YKS alıcısından, alıcı numarası u , toplam ölçüm sayısı N ve örnek numarası n olmak üzere, d günü için kestirilen TEİ değerleri, $\mathbf{x}_{u,d}$ vektörüyle ifade edilir:

$$\mathbf{x}_{u,d} = [\mathbf{x}_{u,d}(1) \dots \dots \mathbf{x}_{u,d}(n) \dots \dots \mathbf{x}_{u,d}(N)]^T \quad (1)$$

Eşitlik (1)'de, T , transpoz operatörü ve $1 \leq n \leq N$ 'dir. Bu çalışmada, KNN sınıflandırıcısının uygulaması için kullanılan parametrelerden biri de IONOLAB-TEC olarak kestirilen TEİ değerleridir [7].

Makine öğrenmesi algoritmaları, parametrik ve parametrik olmayan modeller olarak gruplanabilir. KNN algoritması, örnek tabanlı öğrenme olarak tanımlanan parametrik olmayan modellerin bir alt kategorisine aittir. KNN algoritmasının performansı için kritik öneme sahip noktalardan birisi örnekler arası yakınlığın nasıl ölçümleneceğidir. Yakınlık, Öklid uzaklığı kullanılarak hesaplanabilir [8, 9].

Öklid uzaklığı, sınıflandırma ve kümeleme algoritmalarında en sık kullanılan uzaklık ölçütüdür. Öklid uzaklığı, iki nokta arasındaki doğrusal uzaklık olup, uzayda herhangi iki gündeki iki örnek arasındaki Öklid uzaklığı aşağıdaki gibi tanımlanabilir [8, 9]:

$$d(x_{u;d1}, x_{u;d2}) = \sqrt{\sum_{n=1}^N (x_{u;d1}(n) - x_{u;d2}(n))^2} \quad (2)$$

Öklid uzaklığı, K-Ortalama Kümeleme algoritması, temel KNN algoritması gibi sınıflandırma ve kümeleme algoritmalarında yakınlığın ölçülmesi için kullanılan temel uzaklık ölçütüdür [8].

3. Bulgular

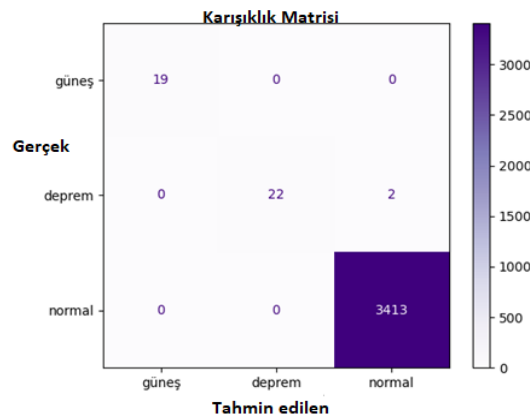
Bu çalışmada, 11 Ağustos 1999 tam güneş tutulmasının ve 17 Ağustos 1999 Marmara depreminin (40.74°K 29.86°D, Mw 7.5, YZ 03.02), iyonkürede meydana getirdiği bozulmaların ve deprem öncülü sinyallerin tespit ve sınıflandırılması için KNN algoritması, IONOLAB-TEC olarak kestirilen TEİ verilerine uygulanmıştır [1, 2, 7]. Çalışma kapsamında 01 ila 20 Ağustos 1999 tarihleri arasındaki yirmi günlük periyot için tubi (40.59° K, 29.45° D), Gebze ve ankr (39.69° K, 32.75° D), Ankara istasyonlarından kestirilen IONOLAB-TEC [7] ve güneş işleklik ve jeomanyetik işleklik indisleri olan Güneş Lekelerinin Sayısı (GLS), güneş akısı F10.7 ve Kp, Ap, AE ve Dst indisleri [10, 11] elde edilmiştir.

Çalışmada, sınıflandırıcı için kullanılan veri kümesine ait tanımlayıcı istatistikler, “NaN” değerleri hariç, bir veri kümesinin dağılımının merkezi eğilimini, dağılımını ve şeklini özetleyenleri içermektedir. Sayısal veriler için, sonucun dizini sayı, ortalama, standart sapma (std), en küçük değer (min), en büyük değer (max) ve alt, 50 ve üst yüzdelik dilimleri içermektedir. Varsayılan olarak, alt yüzdelik dilim 25'tir ve üst yüzdelik dilim 75'tir. 50 yüzdelik dilim, medyan ile aynı elde edilmiştir. Şekil 1’de tüm istatistikler verilmiştir. Şekil 1’de IONOLAB-TEC değerleri “tec” GLS değerleri “sun” ile, F10.7 değerleri de “solar” ile temsil edilmiştir.

	tec	ap	kp	sun	solar	ae	dst
count	11520.00	11520.00	11520.00	11520.00	11520.00	11520.00	11520.00
mean	25.83	14.55	2.73	153.50	124.80	279.80	-9.44
std	9.55	8.18	1.12	31.53	59.75	232.96	15.68
min	7.25	5.00	0.00	123.00	65.00	39.00	-56.00
25%	17.90	8.75	2.00	130.25	74.75	81.00	-20.00
50%	26.09	11.00	3.00	138.00	94.50	203.00	-9.00
75%	32.52	19.75	3.00	171.75	167.25	437.25	3.00
max	51.43	33.00	5.00	216.00	267.00	1155.00	30.00

Şekil 1. IONOLAB-TEC ve güneş ve jeomanyetik işleklik indislerine ait istatistikler.

Toplam 11520 adet örnekten oluşan veri setinin %70’i eğitim, %30’si de test için kullanılmış; bu ayırmda örneklerin seçimi ise rastgele yapılmıştır. Eğitim sürecinden sonra test verileri ile sınıflama başarısı kontrol edilmiştir. Sistemin ürettiği sınıflar ile test sınıfları karşılaştırıldığında ne kadarlık kısmın doğru tahmin edildiği sistemin genel sınıflandırma doğruluğunu göstermektedir. Tüm olası durumların gösterildiği matris olan karışıklık matrisi (Confusion Matrix), Şekil 2’de gösterilmiştir.



Şekil 2. Tüm olası durumlar (deprem, güneş ve normal) için karışıklık matrisi (Confusion Matrix).

Veri kümesinde dört muhtemel sonuç vardır: Gerçek pozitif olan örnek pozitif olarak doğru sınıflandırıldığında Doğru Pozitif (True Positive–TP), gerçek pozitif olan örnek negatif olarak yanlış sınıflandırıldığında Yanlış Negatif (False Negative–FN) olarak nitelendirilir. Ayrıca, gerçek negatif olan örnek negatif olarak doğru sınıflandırıldığında Doğru Negatif (True Negative–TN), gerçek negatif olan örnek pozitif olarak yanlış sınıflandırıldığında Yanlış Pozitif (False Positive–FP) olarak nitelendirilir. Sınıflandırıcının performansı Kesinlik (Precision), Duyarlılık (Recall) ve f1-score (Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını göstermektedir) kullanılarak ölçülmüştür. Doğruluk (Accuracy) yerine f1-score değerinin kullanılmasının en temel

sebebi eşit dağılmayan veri kümelerinde hatalı bir model seçimi yapmamaktır. Şekil 3'te sınıflandırma sonucu elde edilen sonuçlar görülmektedir.

	precision	recall	f1-score	support
gunes	1.00	1.00	1.00	19
deprem	1.00	0.92	0.96	24
normal	1.00	1.00	1.00	3413
accuracy			1.00	3456
macro avg	1.00	0.97	0.99	3456
weighted avg	1.00	1.00	1.00	3456

Şekil 3. Tüm olası durumlar için f1-score ve sınıflandırma oranları.

Çalışmada, “depren” ile ifade edilen olası durum, depremden önceki zaman aralığını; “güneş” ile ifade edilen olası durum da, güneş tutulması zaman aralığını ifade etmektedir. “normal” olarak tanımlanan olası durum da, tutulmanın ve deprenin olmadığı, iyonkürenin sakin olarak nitelendirilen durumunu tanımlamaktadır.

4. Sonuç

Bu çalışmada, 11 Ağustos 1999 tam güneş tutulmasının ve 17 Ağustos 1999 Marmara depreninin iyonkürede meydana getirdiği bozulmaların tespiti ve tahmini için, K-En Yakın Komşular (KNN) algoritması, IONOLAB-TEC verilerine uygulanmıştır. Test setindeki 19 güneş tutulması değerinin tamamı doğru sınıflandırılmıştır. 24 depren verisinin 22 tanesi doğru bulunmuş 2 tanesi ise hatalı olarak sınıflandırılmıştır. Normal diye adlandırılan 3413 test verisinin ise yine tamamı doğru sınıflandırılmıştır.

Kaynaklar

- [1]. Arıkan F., Erol C., ve Arıkan O., "Regularized estimation of vertical total electron content from Global Positioning System data", Space Physics, cilt.108 no.A12, s.1-12, 2003.
- [2]. Arıkan F., Nayir H., Sezen U., ve O. Arıkan, "Estimation of single station interfrequency receiver bias using GPS-TEC", Radio Science, cilt.43, no.4, s.1-13, 2008.
- [3]. Pulnits S., Legen'ka A., Gaivoronskaya T., ve Depue V., "Main phenomenological features of ionospheric precursors of strong earthquakes", Journal of Atmospheric and Solar Terrestrial Physics, cilt.65, no.16-18, s. 1337-1347, 2003.
- [4]. Karatay S., Arıkan F., ve Arıkan O., "Investigation of total electron content variability due to seismic and geomagnetic disturbances in the ionosphere", Radio Science, cilt.45, no.5, s.1-12, 2010.
- [5]. Akyol A., Arıkan O., ve Arıkan F., "A Machine Learning-Based Detection of Earthquake Precursors Using Ionospheric Data", Radio Science, cilt.55, no.11, s.1-21, 2020.
- [6]. Fuying Z., Yun W., Yiyan Z., ve Jian L., "A statistical investigation of pre-earthquake ionospheric TEC anomalies", Geodesy and Geodynamics, cilt.2, no.1, s.61-55, 2011.
- [7]. Ionolab, www.ionolab.org.
- [8]. Taşçı A., ve Onan A., "K En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi", Akademik Bilişim, Aydın, s.1-8, 2016.
- [9]. Raschka S., Python Machine Learning, Birmingham: Packt Publishing, 2015.
- [10]. National Oceanic and Atmospheric Administration, ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/.
- [11]. World Data Center, http://wdc.kugi.kyoto-u.ac.jp/index.html.